

INFORMATION PROCESSING APPARATUS

BACKGROUND OF THE INVENTION

The present invention relates to an information
5 processing apparatus, such as a lockstep fault tolerant computer,
that simultaneously processes the same instructions in a
plurality of clock-synchronized computer modules therein, and
more particularly, to an information processing apparatus that
speedily synchronizes a computer module, which has been out of
10 synchronism with the other computer modules and isolated from
the operation, with other computer modules.

A conventional lockstep fault tolerant computer has a
plurality of computer modules which simultaneously execute the
same instructions. In the fault tolerant computer, one of the
15 computer modules may operate differently from the other computer
modules because of a failure or some other causes. Upon detecting
a computer module that operates differently from the other
computer modules, in other words, on finding a computer module
which is out of lockstep synchronism, the lockstep fault tolerant
20 computer once puts the detected computer module out of the
operation.

Causes which make the computer module be out of the
lockstep synchronism vary. A course of reaction to be taken
for the computer module, which is out of the lockstep synchronism,
25 depends on the cause. One of the causes, which makes the computer
module be out of the lockstep synchronism, may be a permanent
failure that occurs within the computer module. The permanent

failure is not a temporary disturbance or a failure that recovers by the computer module itself, but a failure requiring repairs. A computer module, in which a permanent failure occurs, is usually taken out of the lockstep fault tolerant computer and, instead
5 of that module, another healthy computer module is installed.

Another potential cause, which makes the computer module be out of the lockstep synchronism, may be a lack of synchronism that the operation timing does not synchronize temporarily with the other computer modules because of manufacturing variations
10 of the computer modules. Yet another potential cause may be temporary malfunction of a memory in the computer module affected by an influence such as an α ray. In those causes like a lack of synchronism or temporary malfunction, which does not cause a permanent failure, the computer module need not be replaced.

15 If the permanent failure occurs, the faulty computer module is replaced and the replaced computer module is joined to and synchronized with the other computer modules. If there is no permanent failure, the computer module is rejoined to and resynchronized with the other computer modules. The operation
20 to make a disconnected computer module rejoin the other computer modules is a resynchronization. When the conventional lockstep fault tolerant computer resynchronizes with the computer module which was out of the lockstep synchronism, the conventional lockstep fault tolerant computer copies a memory of the computer
25 module, which is to be rejoined, from a memory of another computer module which is in the lockstep synchronism. The rejoined

computer module thereafter executes the same operations with the other computer modules.

A conventional lockstep fault tolerant computer forces all computing modules stop and copies the whole contents of memory of the joined or rejoined computer module from another computer module being in the lockstep synchronism when joining or rejoining the computing module. This allows all the computing modules to have completely the same internal state. A conventional lockstep fault tolerant computer is forced to stop long time to join or rejoin the computer module. This is because it takes a long time to copy the whole contents of the memory in the computer module. Especially, as memory size in the computer module increases, time to copy the whole content of the memory in the computer module increases.

15

BRIEF SUMMARY OF THE INVENTION

An object of the present invention is to provide an information processing apparatus that ameliorates availability.

20 Another object of the invention is to provide an information processing apparatus that quickly resume operation after the detection of a failure.

According to one aspect of the present invention, an information processing apparatus is provided which includes: first and second computer elements which execute the same instructions substantially simultaneously in substantial synchronism, and which have first and second memory elements,

25

respectively; a monitor element which finds which of the computer elements is out of the synchronism; a copy element which copies a part of the data stored in the second memory element to the first memory element when the monitor element finds that the first computer element is out of the synchronism; and a third memory element which stores information to designate which part of the data stored in the second memory element is copied by the copy element when the monitor element finds that the first computer element is out of the synchronism.

10 According to another aspect of the present invention, an information processing apparatus is provided which includes: first and second computer elements which execute the same instructions substantially simultaneously in substantial synchronism, which have first and second memory elements, respectively, and each of which has at least one processor and a bus connected to the processor; a monitor element which is connected to the bus and which finds which of the computer elements is out of the synchronism; a copy element which copies a part of the data stored in the second memory element to the first memory element when the monitor element finds that the first computer element is out of the synchronism; and a third memory element which stores information to designate which part of the data stored in the second computer element is copied by the copy element when the monitor element finds that the first computer element is out of the synchronism.

BRIEF DESCRIPTION OF THE DRAWINGS

Other features and advantages of the invention will be made more apparent by the following detailed description and the accompanying drawings, wherein:

Fig.1 is a block diagram showing a embodiment of the present invention; and

Fig.2 is a diagram showing an example of operation of the present invention.

In the drawings, the same reference numerals represent the same structural elements.

10

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

As described in the prior art, a cause that puts a computer module out of lockstep synchronism is a permanent failure or a non-permanent failure. In a fault tolerant computer, a computer module in which the permanent failure is occurred must be replaced. On the other hand, if a computer module that is out of the lockstep synchronism because of a non-permanent failure, it is usually not replaced but rejoined unchanged. Namely, in considerable cases, a computer module which is out of the lockstep synchronism is not replaced but installed unchanged. There may be a difference between data stored in a memory of the computer module, which is out of the lockstep synchronism, and data stored in a memory of the computer module, which is in the lockstep synchronism, while the memories of the computer modules will store the same data if no failure is detected. In many cases, the difference is a little or limited.

An embodiment of the present invention will be described in detail below.

Referring to Fig.1, an information processing apparatus includes computer modules 100, 200 and 300, peripheral device
5 controllers 400 and 500, a monitoring element 700, an address storing element 701 and a data transmission element 702. In this embodiment, the information processing apparatus is a lockstep fault tolerant computer.

Computer module 100 includes processors 101 and 102, a bus
10 103, a memory 104 and a memory controller 105. Processors 101 and 102 have the same or an equivalent configuration and are connected to the same bus 103. Memory controller 105 is connected to bus 103. Processors 101 and 102 are connected to memory controller 105 via bus 103. Memory 104 is connected to memory
15 controller 105. Memory controller 105 is connected to data transmission element 702 via a signal line 730. Memory controller 105 is connected to peripheral device controller 400 via a signal line 600 and peripheral device controller 500 via a signal line 610.

20 Every computer modules 100, 200 and 300 has the same or an equivalent configuration or structure. Specifically, computer module 200 includes processors 201 and 202, a bus 203, a memory 204 and a memory controller 205. Processors 201 and 202 are connected to the same bus 203. Memory controller 205
25 is connected to data transmission element 702 via a signal line 731. Memory controller 205 is connected to peripheral device controller 400 via a signal line 601 and peripheral device

controller 500 via a signal line 611. Computer module 300 includes processors 301 and 302, a bus 303, a memory 304 and a memory controller 305. Processors 301 and 302 are connected to the same bus 303. Memory controller 305 is connected to data
5 transmission element 702 via a signal line 732. Memory controller 305 is connected to peripheral device controller 400 via a signal line 602 and peripheral device controller 500 via a signal line 612.

Next, an embodiment of the present invention will be
10 described in more detail below. For concise explanation, the description is focused on computer module 100.

Processors 101 and 102 execute instructions instructed by the lockstep fault tolerant computer 1. The instruction execution by processors 101 and 102 is substantially synchronized
15 with that by the processors of computer modules 200 and 300 based on an identical or substantially the same clock signal, and processors 101 and 102 execute the same or substantially the same instructions substantially simultaneously with the processors of computer modules 200 and 300. The source of the
20 clock signal is provided commonly for the all computer modules 100, 200 and 300, or the sources of the clock signals, which are synchronized, are provided for computer modules 100, 200 and 300, respectively. Namely, computer modules 100, 200 and 300 execute in the instructions "lockstep" synchronism in which
25 every computer modules 100, 200 and 300 execute a substantial identical instruction stream substantially simultaneously. During the instruction execution, processors 101 and 102 write

data into or read data from memory 104. Processors 101 and 102, which is synchronized with the processors of computer modules 200 and 300 based on the clock signal, accesses a peripheral device or peripheral devices. Specifically, processors 101 and 102 access the peripheral device connected to peripheral device controller 400 via bus 103, memory control element 105 and signal line 600. Processors 101 and 102 access the peripheral device connected to peripheral device controller 500 via bus 103, memory control element 105 and signal line 610. When processors 101 and 102 receive an interrupt, which is a stop direction, from monitoring element 700, processors 101 and 102 write context of a process or processes, which is or are executed at the time when the interrupt is received, into the predetermined area of the memory and stop their operation. If processors 101 and 102 stop their operation because of the stop direction arisen from their own reason that they are out of the lockstep synchronism, processors 101 and 102 execute hardware diagnosis afterward. The hardware diagnosis is an execution to diagnose the hardware of computer modules 100 whether or not there is any failure.

Memory controller 105 sends access requests, which are the write access requests and/or the read access requests received from processors 101 and/or 102, to memory 104. Memory controller 105 sends responses from memory 104 to processors 101 and 102. The request is send from processors 101 and 102 to memory 104 when the access request is the write access request or the read access request. The write access request includes write data. The response is send from memory 104 to processors 101 and 102

when the request is the read access request. The response includes read data. The memory controller 105 sends access requests, which are came from processors 101 and/or 102 and are addressed to at least one peripheral device, to peripheral device
5 controllers 400 and 500. The memory controller 105 sends access requests, which are received from data transmission element 702 via signal line 730, to memory 104. For example, the access received from data transmission element 702 is to execute direct memory access (DMA) transmission. In the DMA transmission,
10 memory 104 is either an origin of the transmission or a destination of the transmission.

Peripheral device controllers 400 and 500 monitor whether or not access requests to the peripheral device received from all of computer modules 100, 200 and 300 differ each other. If
15 none of the access requests received from all of computer modules 100, 200 and 300 differs, each of peripheral device controllers 400 and 500 sends a single access request out of the access requests to the corresponding peripheral device. If any of the access requests received from all of computer modules 100, 200 and 300
20 differs from the others, each of peripheral device controllers 400 and 500, for example, discards these access requests or sends a single access request, which is determined by majority decision rule, to the corresponding peripheral device. When the access request addressing to the peripheral device is the read access
25 request, each of peripheral device controllers 400 and 500 send a response, which include data read out from the corresponding

peripheral device, to all of the computer modules 100, 200 and 300 simultaneously.

In this embodiment, monitoring element 700 is connected to a bus which is directly connected to processors 101 and 102.

5 This accelerates the detection by monitoring element 700, which is to find which of computer modules 100, 200 and 300 is out of the lockstep synchronism. Monitoring element 700 is connected to bus 103 of computer module 100 through signal lines 710 and 720. In the access request from processors 101 and 102

10 to memory 104 or the peripheral device, signal line 710 distributes an address strobe, which indicates the time when the address is output, from bus 103 to monitoring element 700. In the access request from processors 101 and 102 to memory 104 or the peripheral device, signal line 720 distributes a command

15 and an address from bus 103 to monitoring element 700. The command includes, for example, a write access command or a read access command. Monitoring element 700 is connected to bus 203 of computer module 200 through signal lines 711 and 721 and is connected to bus 303 of computer module 300 through signal lines

20 712 and 722.

Monitoring element 700 finds which of computer modules 100, 200 and 300 is out of the lockstep synchronism. Monitoring element 700 monitors the consistency of the access requests from computer modules 100, 200 and 300 on the basis of the address

25 strobes received via signal lines 710, 711 and 712 and the commands and the addresses received via signal lines 720, 721 and 722. When monitoring element 700 detects the inconsistency of the

access requests from computer modules 100, 200 and 300, monitoring element 700 notifies address storing element 701 that there is the inconsistency of the access requests between computer modules 100, 200 and 300 and which one is the inconsistent
5 computer module. The computer module whose access request is inconsistent with the other computer modules is determined to be out of the "lockstep" synchronism. When monitoring element 700 detects the inconsistency, monitoring element 700 notifies the processors of all computer modules 100, 200 and 300 of a
10 stop direction, which is in fact an interruption to the processors of computer modules 100, 200 and 300. On receiving the stop direction, each processor writes context of a process or processes prosecuted at the time of the interruption into the predetermined location of the memory, and then halts. In an
15 example of monitoring the consistency of the access requests between computer modules 100, 200 and 300, monitoring element 700 detects the consistency or the inconsistency of the access requests when monitoring element 700 receives the address strobes from every computer modules 100, 200 and 300 during the same
20 cycle, and the commands and the addresses at this cycle are the same between computer modules 100, 200 and 300. If an address of an access request from computer module 100 is different from addresses of access requests from computer modules 200 and 300 during a certain cycle, computer module 100 is found to be out
25 of lockstep synchronism, in other words, inconsistent. In another example, which is a simplified example, monitoring element 700 receives only the address strobes from all computer

modules 100, 200 and 300, and determines the consistency or the inconsistency of the access requests when the address strobes from computer modules 100, 200 and 300 are received during the same cycle.

5 Address storing element 701 has a buffer which stores a address or addresses corresponding to the data, which is stored in the memory of the computer module being in the lockstep synchronism and which differ from the data stored in the memory of the computer modules being out of the lockstep synchronism.

10 Address storing element 701 stores a address or addresses directed by the access request in which the inconsistency is detected and the write access requests afterwards by computer modules 100, 200 and 300, since monitoring element 700 notifies address storing element 701 of the inconsistency of the access

15 request and the inconsistent computer module.

 Data transmission element 702 interrogates an error indicator flag and a hardware diagnosis result, when all processors of computer modules 100, 200 and 300 halt and a hardware diagnosis afterward is completed,. The error indicator flag

20 is a flag which indicates that an error occurred in the computer module. If a permanent failure occurred in the computer module, data transmission element 702 is able to find it out based on the error indicator flag and the hardware diagnosis result. The permanent failure is not a temporary disturbance or a failure

25 that recovers by itself, but a failure requiring repairs. Data transmission element 702 executes a resynchronization if no permanent failure is occurred in the computer module. The

resynchronization includes execution to conform memory contents of the computer module being out of the lockstep synchronism to the memory contents of the other computer modules which are in lockstep synchronism. In the resynchronization, if the
5 computer module has a cache, specifically, if the processors have a cache, cache flash operations are executed in the computer module which is in lockstep synchronism. A cache flash operation may be executed only in a single computer module which is in lockstep synchronism. By the cache flash operations, the data
10 in the cache is written out to the memory. An address or addresses, which correspond to the data from the cache written to the memory, are stored in address storing element 701. After the completion of the cache flash, data transmission element 702 copies the data, which corresponds to the address or the addresses stored
15 in address storing element 701, of the memory of the computer module being out of lockstep synchronism from the memory of the computer module being in lockstep synchronism. Namely, the data, which is designated by the address or the addresses stored in address storing element 701 and which is stored in the memory
20 of the computer module being lockstep synchronism, is copied to the memory of the computer module which is out of lockstep synchronism. In this copy operation, a direct memory access (DMA) transmission may be utilized.

After data transmission element 702 completes the copy
25 operation, data transmission element 702 resets all computer modules 100, 200 and 300 and make them resume the executions. All computer modules 100, 200 and 300 start ordinary execution.

All processors in computer modules 100, 200 and 300 use the context stored in the predetermined memory area of the computer module to start ordinary execution.

In the above described embodiment, signal lines 710 and
5 720, which are derived from bus 103, are used to transmit the access request addressing to memory 104 from processors 101 and 102 to monitoring element 700 and address storing element 701. In a restricted case, the present invention may be modified. For example, this modification is to use signal lines derived
10 from the line, which connects memory controller 105 and memory 104, to transmit the access request from processors 101 and 102 to monitoring element 700 and address storing element 701.

Next, the operation of this embodiment will be described.

Referring to Figs. 1 and 2, computer modules 100, 200 and
15 300 ordinarily execute operations in the lockstep synchronism. Namely, computer modules 100, 200 and 300 ordinarily execute the same instructions substantially simultaneously based on an identical or substantially the same clock signal. The processors of computer modules 100, 200 and 300 access the memory
20 and the peripheral device in accordance with the instructions. Monitoring element 700 monitors every access from computer modules 100, 200 and 300. Specifically, monitoring element 700 watches the time, the command and the address of the access requests in the same cycle whether they are consistent between
25 computer modules 100, 200 and 300.

Assuming that computer module 100 is disturbed and thus the access request from computer module 100 is inconsistent with

the access requests from the other computer module 200 and 300 but no permanent failure occurred in computer module 100, monitoring element 700 detects the inconsistent. On detecting the inconsistent, monitoring element 700 determines which
5 computer modules 100, 200 and 300 is out of lockstep synchronism. In this embodiment, monitoring element 700 determines that computer module 100 is out of lockstep synchronism. Monitoring element 700 notifies address storing element 701 of the access inconsistency and the computer module being out of the lockstep
10 synchronism, in this embodiment, computer module 100. Monitoring element 700 notifies all the processors in computer modules 100, 200 and 300 of the stop direction by interruption.

When address storing element 701 is notified of the access inconsistency and the computer module 100 being out of the
15 lockstep synchronism, address storing element 701 records the addresses of the inconsistent access request and the write access requests thereafter from each of computer modules 100, 200 and 300.

The processor, which is notified of the stop direction,
20 writes a context of an ongoing process or ongoing processes to the predetermined area of the memory and then halts. The hardware diagnosis is executed on the computer module whose access is inconsistent with the other computer module. In this example, the hardware diagnosis is executed on computer module 100. After
25 the completion of the hardware diagnosis, data transmission element 702 interrogates the error indicator and the hardware diagnosis result. Since no permanent failure is occurred in

computer module 100 in this embodiment, data transmission element 702 executes the resynchronization.

In the resynchronization, if any computer module being in lockstep synchronism has the cache, the cache flash is executed.

5 The cache flash is executed in, for example, computer module 200. In this embodiment, cache flash is to read out the whole contents of the cache to an area of the memory of the computer module. The cache flash makes the data in the cache be written out to the memory. This written out operation to the memory
10 is executed by the write access, and the address whose data is written out is stored in address storing element 701.

Data transmission element 702 copies the data, which corresponds to the address or the addresses stored in address storing element 701 only and which is stored in the memory of
15 one of the other computer modules, which are in the lockstep synchronism, computer module 200 in this embodiment, to the memory of the computer module to be resynchronize, computer module 100 in this embodiment. In this embodiment, the copy operation utilizes the DMA transmission. The number of the
20 addresses stored in address storing element 701 is less than the number of the entire addresses of the memory. The copy of the data in the present invention based on the addresses stored in address storing element 701 takes less time than the copy of the data of the entire addresses. After the completion of
25 the copy operation, data transmission element 702 resets all computer modules 100, 200 and 300. Subsequent to the reset, all computer modules 100, 200 and 300 is synchronized with the

identical or the substantially the same clock signal each other,
and start the ordinary execution.

As described above, when monitoring element 700 finds that
any computer module is out of the lockstep synchronism, address
5 storing element 701 stores the address or addresses, whose data
has possibility to differ from the corresponding data of the
other computer modules, of the memory of the computer module
which is out of lockstep synchronism. And, during
resynchronization, data transmission element 702 copies the data
10 corresponding to the address or addresses stored in address
storing element 701 from the memory of the computer module which
is in the lockstep synchronism to the memory of the computer
module being out of lockstep synchronism. The time to complete
the copy of the memory of the resynchronizing computer module
15 is shortened. As a result, it is possible to mount the computer
module, which was once out of the lockstep synchronism because
of no crucial reason such as the permanent failure, into the
fault tolerant computer as early as possible.

In this embodiment, for a purpose of explanation, three
20 computer modules 100, 200 and 300 are provided in the lockstep
fault tolerant computer 1. The present invention is not limited
to such a particular configuration, and the number of the computer
modules may be no less than two.

While this invention has been described in conjunction with
25 the preferred embodiments described above, it will now be
possible for those skilled in the art to put this invention into
practice in various other manners.